

(12) DEMANDE INTERNATIONALE PUBLIÉE EN VERTU DU TRAITÉ DE COOPÉRATION
EN MATIÈRE DE BREVETS (PCT)

(19) Organisation Mondiale de la Propriété
Intellectuelle
Bureau international



(43) Date de la publication internationale
27 octobre 2005 (27.10.2005)

PCT

(10) Numéro de publication internationale
WO 2005/101292 A2

(51) Classification internationale des brevets⁷ : G06K 9/00

(74) Mandataires : HASSINE, Albert etc.; Cabinet Plasseraud, 65/67, rue de la Victoire, F-75440 Paris Cedex 09 (FR).

(21) Numéro de la demande internationale :

PCT/FR2005/000673

(22) Date de dépôt international : 18 mars 2005 (18.03.2005)

(25) Langue de dépôt : français

(26) Langue de publication : français

(30) Données relatives à la priorité :

0403556 5 avril 2004 (05 04 2004) FR

(71) Déposant et

(72) Inventeur : LEBRAT, François [FR/FR]; 98 avenue de Versailles, F-75016 Paris (FR)

(81) États désignés (sauf indication contraire, pour tout titre de protection nationale disponible) : AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW

[Suite sur la page suivante]

(54) Title: METHOD FOR SEARCHING CONTENT PARTICULARLY FOR EXTRACTS COMMON TO TWO COMPUTER FILES

(54) Titre : PROCÉDÉ DE RECHERCHE DE CONTENU, NOTAMMENT D'EXTRAITS COMMUNS ENTRE DEUX FICHIERS INFORMATIQUES

AA			L	a		t	o	r	t	u	e	
AD1		SN1	0		1		2		3		4	AD2
BB			0		?		1		1		0	SN2
L	0	0										
e												
	1	0										
i												
i	2	?										
è												
v	3	1										
r												
e	4	0										

AA THE TORTOISE

BB THE HARE

(57) Abstract: The present invention relates to searching content, particularly for at least one extract common to a first data file and a second data file. The method comprises a preliminary step of preparing at least the first file by (a) dividing the first file into a series of data packets having a predetermined size, and identifying packet addresses in said file, (b) combining each packet address with a digital signature that defines one of three fuzzy logic states, namely true, false and indeterminate, and is the result of a combinatorial computation on data from said file; whereafter said method comprises performing an actual search for a common extract by (c) comparing the fuzzy logic states combined with each packet address of the first file with fuzzy logic states determined on the basis of data from the second file, and (d) removing from said common extract search the respective address pairs from the first and second files that have the respective logic states true and false

or false and true, and retaining the other address pairs that identify data packets that may comprise said common extract.

[Suite sur la page suivante]

WO 2005/101292 A2



(84) États désignés (sauf indication contraire, pour tout titre de protection régionale disponible) : ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), eurasién (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), européen (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)

Publiée :

— sans rapport de recherche internationale, sera republiée dès réception de ce rapport

En ce qui concerne les codes à deux lettres et autres abréviations, se référer aux "Notes explicatives relatives aux codes et abréviations" figurant au début de chaque numéro ordinaire de la Gazette du PCT.

Déclaration en vertu de la règle 4.17 :

— relative à la qualité d'inventeur (règle 4.17 iv)) pour US seulement

(57) Abrégé : L'invention concerne la recherche de contenu, notamment une recherche d'au moins un extrait commun entre un premier fichier et un second fichier de données. Le procédé de l'invention comporte une préparation préalable du premier fichier au moins, comprenant les étapes suivantes : a) segmenter le premier fichier en une succession de paquets de données, de taille choisie, et identifier des adresses de paquets dans ledit fichier, b) associer à l'adresse de chaque paquet une signature numérique définissant un état en logique floue parmi au moins trois états : "vrai", "faux" et "indéterminé", ladite signature résultant d'un calcul combinatoire sur des données issues dudit fichier, et le procédé se poursuit par une recherche d'extrait commun, proprement dite, comprenant les étapes suivantes, c) comparer les états de logique floue associés à chaque adresse de paquet du premier fichier, avec des états de logique floue déterminés à partir de données issues du second fichier, d) éliminer de ladite recherche d'extrait commun des couples d'adresses respectives des premier et second fichiers dont les états logiques respectifs sont "vrai" et "faux" ou "faux" et "vrai", et conserver les autres couples d'adresses identifiant des paquets de données susceptibles de comporter ledit extrait commun